Amendments to the Claims

The listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

Claim 1 (currently amended): An integrated human and computer interactive data mining method comprises the steps of:

    a) Input a database;

    b) Perform learning, modeling and analysis using the database to create an initial knowledge model wherein the initial knowledge model is an enhanced regulation tree deriving and storing additional statistics at each node allowing incremental update of rules and multi-level abstraction visualization;

    c) Perform a query of the initial knowledge model using a query request;

    d) c) Perform visualization processing of the initial knowledge model to create a knowledge presentation output containing

            i. Ranks for the rules associated with each of the tree terminal nodes,

            ii. Contrast examples for each terminal node,

            iii. Associated feature distribution profiles for each non-terminal node.

Claim2 (canceled)

Claim 3 (currently amended): The integrated human and computer interactive data mining method of claim 1 wherein the knowledge presentation output further comprises rule ranking by information integration using global characteristics and population characteristics selected from the set consisting of:

    a) Local counts confidence for class c in a terminal node n is defined as:

$$LC_c^{\,n} = \frac{N_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} N_c^{\,n}};$$

b) Local population ~~statisties~~ confidence for class c in a terminal node n is defined as:

$$LP_c^{\,n} = \frac{P_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} P_c^{\,n}};$$

c) Global counts confidence for class c in a terminal node n is defined as:

$$GC_c^{\,n} = \frac{G_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} G_c^{\,n}};$$

d) Global population ~~statisties~~ confidence for class c in a terminal node n is defined as:

$$GP_c^{\,n} = \frac{g_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} g_c^{\,n}}.$$

Claim 4 (currently amended): The integrated human and computer interactive data mining method of claim 1 wherein the knowledge presentation output includes feature distribution profiles wherein a feature is normalized and weighted ranked and the ranks of the samples for the feature are prepared for quick feature visualization.

Claims 5-6 (canceled)

Claim 7 (currently amended): An integrated human and computer interactive data mining method comprises the steps of:
   a) Input a database;
   b) Perform knowledge creation selected from the group consisting of learning, modeling, and analysis using the database to create an initial knowledge

model wherein the initial knowledge model is an enhanced regulation tree deriving and storing additional statistics at each node allowing incremental update of rules;

c) Perform dynamic learning and knowledge representation using the initial knowledge model and the database to create or update a presentable knowledge model facilitating interaction with human by rule prioritization, clustering, and contrast example selection wherein contrast examples are selected from samples of similar characteristics that match the selected rule yet have distinctively different labels.;

Claims 8-13 (canceled)

Claim 14 (currently amended): A multiple level integrated human and computer interactive data mining method comprises the steps of:

a) Input a database;

b) Perform knowledge creation selected from the group consisting of learning, modeling, and analysis using the database to create an initial knowledge model wherein the initial knowledge model is an enhanced regulation tree deriving and storing additional statistics at each node allowing incremental update of rules;

c) Perform overview interactive data mining and dynamic learning and knowledge representation using the initial knowledge model and the database to create or update a presentable knowledge model wherein overview uses a shallower enhanced regulation tree created by pruning the deep tree or simply limiting the depth of the tree and combining all nodes beyond the depth limit.

Claim 15 (currently amended): The multiple level integrated human and computer interactive data mining method of claim 14 further performs zoom and filter interactive data mining and dynamic learning and knowledge representation using the presentable knowledge model and the database to create or update the presentable knowledge model

<u>wherein zoom and filter allow certain branches of the enhanced regulation tree to expand following user direction.</u>

Claims 16-17 (canceled)

Claim 18 (currently amended): A presentable knowledge model generation method comprises the steps of:

 a) Input formatted data and a decision tree;

 b) Perform rule ranking using the formatted data and the decision tree to create ranked output;

 c) Perform feature profile generation using the formatted data and the decision tree to create feature profiles <u>wherein a feature is normalized and weighted ranked and the ranks of the samples for the feature are prepared for quick feature visualization.</u>

 d) Group the ranks and feature profiles to create a presentable knowledge model output.

Claim 19 (currently amended): The presentable knowledge model generation method of claim 18 further comprising perform contrast example selection using the formatted data and the decision tree to create contrast examples and group the contrast examples, ranks, and feature profiles to create a presentable knowledge model output <u>wherein contrast examples are selected from samples of similar characteristics that match the selected rule yet have distinctively different labels.</u>

Claim 20 (currently amended): The presentable knowledge model generation method of claim 18 wherein rule ranking uses global characteristics and population characteristics selected from the set consisting of:

 a) Local counts <u>confidence for class c in a terminal node n is defined as:</u>

$$LC_c^{\,n} = \frac{N_c^{\,n}}{\displaystyle\sum_{c \in All\_Classes\_in\_n} N_c^{\,n}} \;;$$

b) Local population ~~statistics~~ confidence for class c in a terminal node n is defined as:

$$LP_c^{\,n} = \frac{P_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} P_c^{\,n}} \; ;$$

c) Global counts confidence for class c in a terminal node n is defined as:

$$GC_c^{\,n} = \frac{G_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} G_c^{\,n}} \; ;$$

d) Global population ~~statistics~~ confidence for class c in a terminal node n is defined as:

$$GP_c^{\,n} = \frac{g_c^{\,n}}{\sum\limits_{c \in All\_Classes\_in\_n} g_c^{\,n}} \; .$$

Claim 21 (canceled)

Claim 22 (currently amended): The presentable knowledge model generation method of claim 18 wherein the feature profile generation method normalizes the automatically generated features and normalized features are weighted ranked and the ranks of the samples for each feature are prepared for quick feature visualization wherein the weighted ranked process ranks a sample and uses its weight to generate a feature histogram.

Claim 23 (currently amended): A tree update learning method comprises the steps of:
   a) Input formatted data and a decision tree;
   b) Remove a~~sample~~ rule by updating the samples of the rule associated terminal node yet does not update the samples of the rule associated non-terminal node.~~:~~
      i. ~~Subtracting the data from the sample associated terminal node;~~

ii.   Updating the statistics of each of the associated non-terminal nodes.

Claims 24-25 (canceled)

Claim 26 (currently amended): The tree update learning method of claim 23 further comprises a step to add a rule wherein an artificial sample with the feature values match the rule is created and is given high weight subject to the decision tree decision.

Claim 27 (currently amended): An interactive data mining method comprises the steps of:

    a) Input a presentable knowledge model wherein the presentable knowledge model is an enhanced regulation tree deriving and storing additional statistics at each node;

    b) Perform rule viewing of a terminal node in the presentable knowledge model by a parallel coordinate visualization technique that maps a multiple dimensional space onto two display dimensions and at least one data item is presented as a polygonal line wherein the path from the root node through non-terminal nodes toward a terminal node is presented and the weight of a sample is highlighted by the thickness of the line.

Claims 28-29 (canceled)

Claim 30 (currently amended): The interactive data mining method of claim 27 performs data mining using multi-level abstraction wherein a regulation tree with sufficient depth is created as the baseline knowledge model. In the overview mode, a shallower tree is used that could be created by pruning the deep tree or simply limiting the depth of the tree and combining all nodes beyond the depth limit. In the zoom and filter mode, certain branches of the trees could be allowed to expand following user direction. In the details-on-demand mode, the deepest branch of the tree that corresponds to human interest is shown.